TELECOM CHURN CASE STUDY

BY- KOMAL SHARMA
ANIL KUMAWAT
KEERTHANA KOLUM

Business Problem Overview

In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.

For many incumbent operators, retaining high profitable customers is the number one business goal.

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.

In this project, we will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

Explanation of the Dataset Used in the Project

- The dataset used in this project is customer level data from a leading telecom firm.
- It contains information about individual customers and their interactions with the telecom services.
- The dataset includes a variety of variables that capture customer behavior, usage patterns, and demographics.
- These variables provide valuable insights into customer preferences, engagement, and potential indicators of churn.
- By analyzing this dataset, we can uncover patterns and trends that help predict which customers are at high risk of churn

Churn Definition

Definition of churn in the telecom industry: Churn refers to the phenomenon of customers switching from one telecom operator to another. It is a crucial metric in the industry as it indicates the rate at which customers are leaving a service provider.

Explanation of the usagebased definition used in the project: In this project, churn is defined based on customer usage behavior. For prepaid customers, churn occurs when they have not utilized any services, such as outgoing calls, SMS, or mobile internet, over a specific period of time. This definition allows for identifying customers who have stopped using the services and are at risk of switching to another operator.

Differentiating churn in postpaid and prepaid models: Churn is easier to determine in postpaid models, as customers typically inform the operator when they want to terminate their services. In contrast, churn in prepaid models is more challenging to identify, as customers can simply stop using the services without notice. The project focuses on churn prediction for prepaid customers, considering the unique characteristics and challenges associated with this model.

Project Goals

Objectives: Analyze telecom customer data, predict churn, and derive business insights.

Analysis: Comprehensive data analysis to understand customer behavior and preferences.

Prediction: Develop models to forecast customer churn and identify highrisk customers.

Insights: Provide actionable insights to reduce churn rates and improve customer retention.

Data Preprocessing: Clean, handle missing values, remove duplicates, and standardize data formats.

Feature Engineering: Transform raw data into meaningful features to enhance model performance.

Churn Prediction Models

Churn Prediction Models: Built predictive models using machine learning algorithms to identify customers at high risk of churn.

Model Description: Employed advanced techniques such as Random Forest and Logistic Regression to capture complex relationships in the data

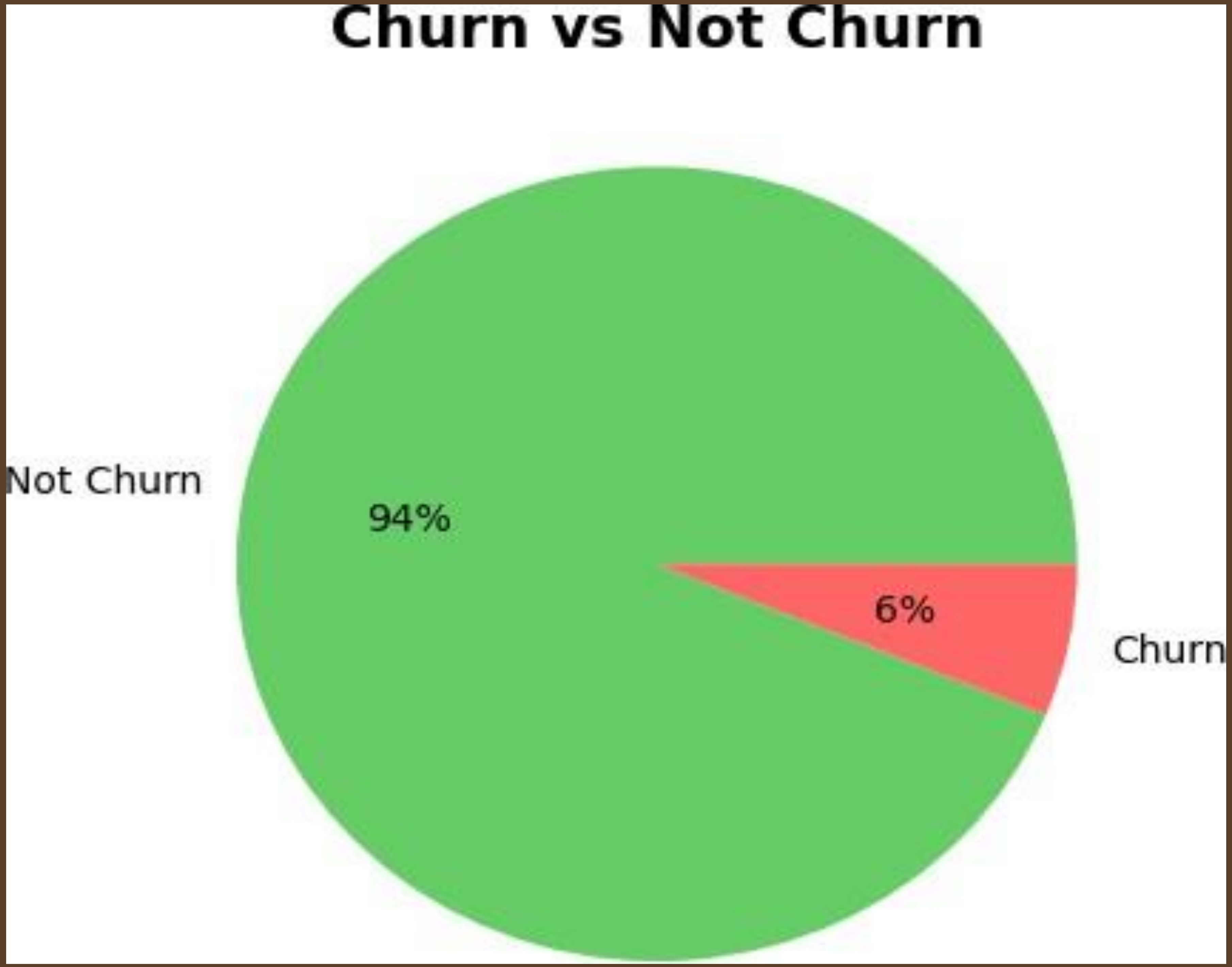
Evaluation Metrics: Assessed model performance using various metrics including accuracy, sensitivity, and specificity.

Performance Assessment: Conducted rigorous evaluation to measure the effectiveness of the models in predicting churn accurately.

Insights and Recommendations: Derived insights from the model results to provide actionable recommendations for reducing churn and improving customer retention strategies.

Results and Insights

TOTAL CUSTOMER CHURN DETAILS



Business recommendation

Top predictors

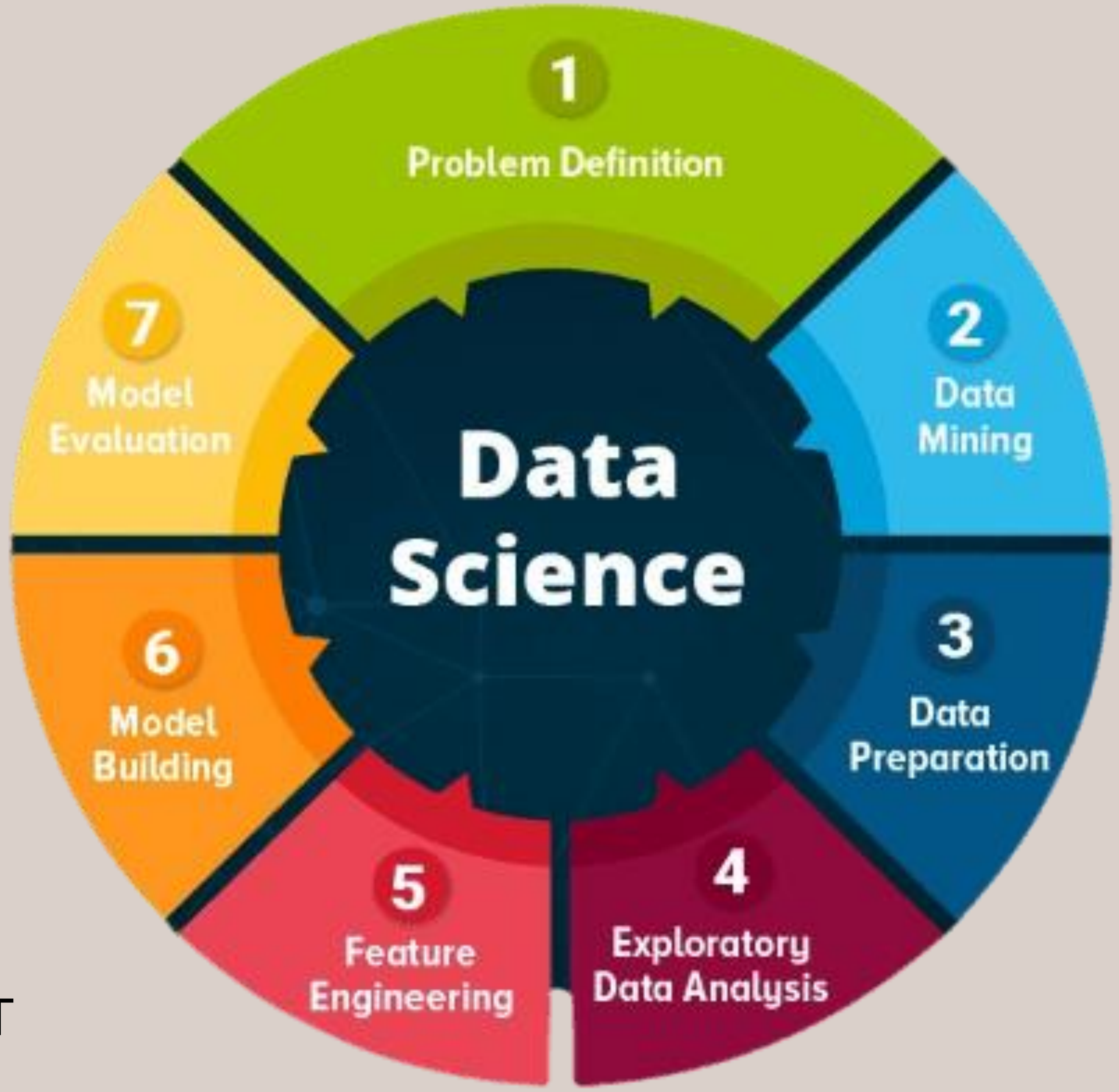
Below are few top variables selected in the logistic regression model.

We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probablity.
E.g.:-
If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

Variables	Coefficients
loc_ic_mou_8	-3.3287
og_others_7	-2.4711
ic_others_8	-1.5131
isd_og_mou_8	-1.3811
decrease_vbc_action	-1.3293
monthly_3g_8	-1.0943
std_ic_t2f_mou_8	-0.9503
monthly_2g_8	-0.9279
loc_ic_t2f_mou_8	-0.7102
roam_og_mou_8	0.7135

Recommendations

1.Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
2.Target the customers, whose outgoing others charge in July and incoming others on August are less.
3.Also, the customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.
4.Cutomers, whose monthly 3G recharge in August is more, are likely to be churned.
5.Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
6.Cutomers decreasing monthly 2g usage for August are most probable to churn.
7.Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
8.roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.



THANK YOU